# Multiple Description Coding and Path Diversity for Voice Communication over MANETs

Jagadeesh Balam and Jerry D. Gibson
Department of Electrical and Computer Engineering
University of California, Santa Barbara, California 93106-9560
Email: balam@ece.ucsb.edu and gibson@ece.ucsb.edu

***Abstract*— We compare a single description coder (G.729) over a single path (SD), a new multiple description coder based on G.729 (MD-G.729) with path diversity (MD-PD) and a duplicated full-rate single description coder (G.729) with path diversity (DSD-PD) under various packet loss conditions for voice communication over wireless Mobile Ad-hoc Networks (MANETs). We show that for low bitrate speech codecs, using a multiple description coder is not very advantageous because the large packet overheads overshadow the small bandwidth savings. Instead, we can use simple path diversity wherein the full rate single description codec is duplicated over independent paths (DSD-PD). Such a method requires only a slightly higher bandwidth than MD-PD but the quality of speech delivered is significantly better when compared to MD-PD. We compare the three different communication methods under random and bursty packet loss conditions on the basis of the quality of the delivered speech. We evaluate the delivered speech quality using the objective speech quality measurement algorithm, PESQ.**

## I. Introduction

Mobile Ad-hoc Networks (MANETs) are formed by mobile wireless hosts without the need of an existing infrastructure, unlike wireless cellular systems which require a centralized control and support system at the base station. Most of the wireless systems deployed today are also centralized systems, wherein the nodes connected to the network communicate through an access point. Interactive voice communication over a wireless mobile ad-hoc network is a challenging problem because of the error prone wireless channel, the changing topology of the network, delays involved in establishing a new link or finding a new route, and the current MAC protocols which were not developed for real-time multimedia communication. One important example is ad-hoc networks based on the IEEE 802.11 standard.

This work was supported by the California Micro Program, Applied Signal Technology, Dolby Labs, Inc. and Qualcomm, Inc., by NSF Grant Nos. CCF-0429884 and CNS-0435527, and by the UC Discovery Grant Program and Nokia, Inc.


The IEEE 802.11 standard specifies a MAC (Medium Access Control) layer protocol and different physical layer protocols like a, b and g operating at different bandwidths and bit rates. The IEEE 802.11 standard was designed primarily for non-real-time transfer of data and these protocols may not be suitable for real time interactive multimedia. The IEEE 802.11 MAC protocols are designed to minimize collisions and depend on retransmissions to ensure successful transmission of a packet irrespective of the delay incurred by the packet. For good quality conversational voice communication, end to end delay of packets must be under 150 ms for the delay to be imperceptible to the listener. Most of the prior work in this area has been on changes that can be made at the MAC layer to minimize delays due to retransmissions and reduce packet losses due to bit errors.

### A. Prior Work

The 802.11 MAC layer retransmits a packet until the packet is acknowledged by the receiver. Retransmissions increase delay and also cause congestion in the network. Many schemes to reduce retransmission of speech packets have been proposed. In [1], the authors suggest selective error checking (SEC), wherein, errors are allowed in some parts of the speech packet and a packet is only dropped if some critical bits in the voice data or the protocol headers are in error. A similar scheme was also mentioned in [2] for an Adaptive Multirate (AMR) coder where a packet is dropped only when there are errors in perceptually important bits. These schemes not only reduce the average delay that each packet undergoes in the network but the overall speech quality is also improved because of a reduced number of packets discarded at the MAC layer after reaching the retransmissions limit. Petracca et al. [3] suggest using forward error correction for only perceptually important packets. The perceptually important packets are determined by computing an analysis-by-synthesis distortion for different parameters in an encoded voice frame. In [4], a variable bit-rate codec is used to to ad apt the bit rate according to instantaneous channel conditions.

### B. Path Diversity and Multiple description coding

Another method to improve reliability of transmission over a MANET is to use path diversity, i.e. send data simultaneously through multiple paths. The probability of all the paths breaking down simultaneously is low and hence the probability of packet loss is reduced, but, sending multiple copies of the same packet is inefficient usage of bandwidth. To improve bandwidth efficiency, a source coding diversity scheme like Multiple Description (MD) coding can be used with path diversity. In multiple description coding, multiple descriptions / bitstreams of the source are created in such a way that each description can be used to reconstruct the source with acceptable quality. When two or more descriptions are available at the receiver, they can be combined to produce a higher quality reconstruction of the source. Using a multiple description coder for voice communication over MANETs was first suggested in [5], where the authors proposed a new MD codec based on the AMR-WB codec and showed that at high error rates in the channel, the MD codec performs better than a single description (SD) codec sent over a single path.

We propose a new MD codec for narrowband speech with balanced side descriptions, based on the G.729 codec. The side descriptions here are of the same average rate and the speech delivered by each description is of similar quality. We compare three different communication methods, 1) using a single description coder (G.729) with a single path (SD), 2) a multiple description coder based on G.729 (MD-G.729) with path diversity (MD-PD), and 3) a duplicated single description coder (G.729) with path diversity (DSD-PD) under random and bursty packet loss conditions. We compare their performance first in a classical situation, where no packet headers are added to each packet. Next, we consider a more practical scenario where packet headers are added to each packet by the various protocol layers in a typical MANET. These headers are typically much larger than the speech payloads and therefore significantly increase the packet loss rate in random bit error channels.

## II. A Multiple Description Speech Coder based on G.729

We designed a new multiple description coder based on the G.729 speech codec. Our MD coder creates two balanced descriptions, i.e. each description is of the same rate, and speech decoded from either description is of similar quality. Such a codec is more suitable for an ad-hoc network, because in a MANET, we cannot guarantee delivery or a better QoS for any one path. The idea behind the coder is to take an SD coder (G.729) and split the bitstream into two sub-streams. This is similar to the no-excess joint rate case of MD coding, where the individual descriptions can be combined to give an optimal joint description. Since dividing the bitstream into two non-overlapping portions cannot give us acceptable quality at the side decoders, we inject some redundancy by replicating vital information in both the descriptions. The distortion at the central decoder is still the same as the SD decoder but the effective bit-rate is higher due to the redundancy introduced in the side descriptions. Of course, the quality delivered by each description will be worse than that of an SD codec optimized for the same rate as an individual description.

*1) Encoder:* ITU-T G.729 is an CS-ACELP based codec for encoding narrowband speech at the rate of 8 kbps. The

TABLE I
BIT ALLOCATION FOR DESCRIPTION I OF MD-G.729

| | Odd Frame | | Even Frame | | Sum |
|---|---|---|---|---|---|
| Frame Indicator | 2(00) | | 2 (01) | | 4 |
| LSP | Stage 1: ***8*** | | Stage 1: ***8*** | | 26 |
| | Stage 2: 5 0 | | Stage 2: 0 5 | | |
| | sf 1 | sf 2 | sf 1 | sf 2 | |
| Pitch delay | 9 | 5 | 0 | 0 | 14 |
| Fixed Codebook | 13 | 0 | 13 | 0 | 26 |
| Fixed Codebook Signs | 4 | 0 | 4 | 0 | 8 |
| Gains | 7 | 0 | 7 | 0 | 14 |
| Total | | | | | 92 |

TABLE II
BIT ALLOCATION FOR DESCRIPTION II OF MD-G.729

| | Odd Frame | | Even Frame | | Sum |
|---|---|---|---|---|---|
| Frame Indicator | 2(10) | | 2 (11) | | 4 |
| LSP | Stage 1: ***8*** | | Stage 1: ***8*** | | 26 |
| | Stage 2: 0 5 | | Stage 2: 5 0 | | |
| | sf 1 | sf 2 | sf 1 | sf 2 | |
| Pitch delay | 0 | 0 | 9 | 5 | 14 |
| Fixed Codebook | 0 | 13 | 0 | 13 | 26 |
| Fixed Codebook Signs | 0 | 4 | 0 | 4 | 8 |
| Gains | 0 | 7 | 0 | 7 | 14 |
| Total | | | | | 92 |

G.729 codec encodes 10 ms speech frames using 80 bits at a resultant bit rate of 8 kbps. The encoder of the MD codec divides the G.729 bitstream into two overlapping bitstreams. Tables I and II show the bit allocations for odd and even frames in each of the descriptions. To keep the effective average bit rate of each description the same (4.6 kbps), odd and even numbered frames in each description are coded with a different number of bits. This is achieved by including the bits corresponding to the pitch delay only in alternate frames. The pitch delay for the second subframe in each frame is differentially encoded with respect to the first subframe. Without the first subframe pitch delay, the second subframe pitch delay cannot be decoded. Hence, pitch delay information for both the subframes has to be always included together in one description. For description I, the 14 bits for adaptive-codebook delay are included in odd-numbered frames and for description II, they are included in even-numbered frames.

Each description has 13 bits allocated to the Line Spectrum Pairs (LSPs). G.729 uses multi-stage split vector quantization to quantize the LSP vector. In the first stage, the vector is not split and 8 bits are used to code the vector. These 8 bits are included in both the descriptions for all the frames. This allows for a coarse reconstruction of the 10-dimensional residual vector of LSPs in either description. In the second stage of the vector quantizer, the 10-dimensional residual vector is split into two 5-dimensional sub-vectors and each sub-vector is coded using 5 bits. For odd (even) numbered frames, the codebook index for the first (second) subvector is included only in description I while the codebook index for the second (first) subvector is included only in description II. This is done to make the descriptions more symmetric with respect to quality. Experiments revealed that the degradation in the reconstructed speech was more when the first subvector is removed rather than when the second sub-vector was removed.

The bits corresponding to the fixed codebook vector and signs of the fixed codebook of the first subframe of all the frames are included only in description I and the same information for the second subframe is included only in description II. The adaptive codebook and the fixed codebook gain information for the first (second) subframe is included only in description I (II). Thus, each odd numbered frame for description I gets 51 bits from the G.729 bitstream while description II gets 37 bits. Similarly, for even numbered frames description I contains 37 bits and description II contains 51 bits. Two frame indicator bits are added to indicate the description to which the bitstream belongs and whether the frame is odd or even numbered. Bit pair '00' indicates that the bitstream belongs to an odd numbered frame of description I, '01' indicates description I and even frame, '10' indicates description II and odd frame and '11' indicates description II and even frame.

*2) Decoder:* When both the descriptions are received at the decoder the two descriptions are combined to give the bitstream of G.729. If both the descriptions are lost, then the frame error concealment algorithm of G.729 [6] is used to conceal the lost frame. If one of the descriptions is received then the decoder substitutes the missing information by using the received parameters in the description or information from the most recent correctly received frame. When only one of the descriptions is received, the LSP vectors are constructed from the received first stage vector and one of the received subvectors. The missing second stage subvector is assumed to be zero. The pitch delay in even (odd) frame in first (second) description is constructed from the previous received frame's pitch delay increased by 1. This process is same as that used for frame error concealment in the G.729 codec. The missing gain information in the second subframe for description I and first subframe for description II is substituted by an attenuated version of the previous subframe. The memory of the gain predictor is also attenuated in a manner similar to that used in G.729 error concealment.

## III. Experiments and Results

For our experiments, we assume that two independent paths with similar channel conditions are always available between the sender and the receiver. We consider two different packet loss conditions: 1) the packet size dependent random packet loss conditions, and 2) bursty packet loss conditions. Under these packet loss conditions, we compare the quality of speech delivered by the three different communication methods mentioned earlier, 1) a single description coder (G.729) over a single path (SD), 2) a multiple description coder with path diversity (MD-PD) and 3) a duplicated full-rate single description coder (G.729) with path diversity (DSD-PD). We first compare the performance of each of the above methods in the classical situation, where no packet headers are added to each packet, and then we investigate the effect of typical packet headers on the performance of each of the above methods in random and burst packet loss conditions.

The quality of the decoded speech is evaluated using PESQ (Perceptual Evaluation of Speech Quality). PESQ, an ITU standard for objective speech quality measurement of narrowband speech, compares the degraded signal with the reference signal and produces a score between -0.5 and 4.5. PESQ scores have been found to correlate well with subjective MOS scores. PESQ-LQ (Listening Quality) was then shown to be a good predictor of subjective listening quality in [7]. PESQ-LQ provides a mapping function to map the PESQ scores to an average ITU-T P.800 MOS (Mean Opinion Score) scale. The mapping function is given by

$$y = \begin{cases} 1.0, x \leq 1.7 \\ -0.157268x^3 + 1.386609x^2 - 2.504699x + \\ 2.0233454, x > 1.7 \end{cases} \quad (1)$$

where $x$ is the PESQ score and $y$ is the corresponding mapping to LQ MOS. We use six different (3 male, 3 female) speech files in our experiments. Each file is about 8 seconds long and consists of two different sentences spoken by a different speaker.

### A. Packetization

We assume that each packet sent over the network contains one 10 ms frame of speech. In an ad-hoc network large delays can occur at each intermediate node because of various factors like contention for the channel or link failure. To allow for the unpredictable delays in the network, we keep the packetization delay at the minimum of 10 ms. Also having more frames per packet impairs the performance of the packet loss concealment algorithm since one lost frame is more effectively concealed than two or more successive lost frames.

### B. Random packet losses

We consider random packet losses that occur due to random bit errors in the channel. For each bit error rate (BER) considered, we first find packet loss probability $p$ using Eq. (2) for the packet length of each codec. For each $p$, 250 different trace files were created using different seeds for the random number generator and frames corresponding to the lost packets in the trace files were dropped in the encoded speech files. The packet loss probability for a given BER is given by,

$$p = 1 - (1 - BER)^L \quad (2)$$

*1) No Packet Headers:* First we consider the classical situation where no packet headers are added to the source information. The SD and DSD-PD methods transmit 80 bits per packet as they use the G.729 codec, while the MD-PD method sends either 53 or 39 bits per packet. For a given BER, it is obvious that SD and DSD-PD have a higher packet loss probability than MD-PD because of the larger packet size.

Figure 1 shows the speech quality delivered by each of the methods for increasing BERs. The better performance of MD-PD compared to SD here can be attributed to the lower

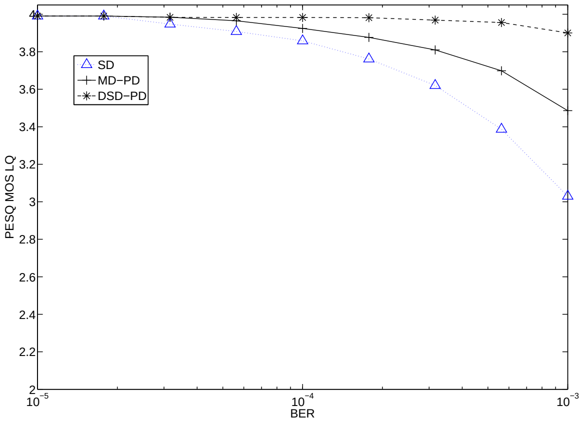


Fig. 1. Average PESQ-MOS LQ for changing BER (without packet headers)

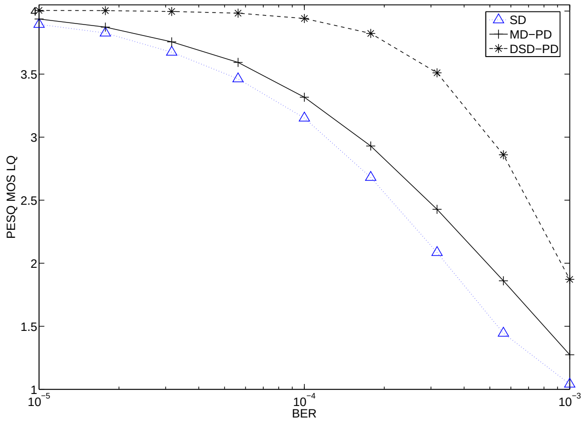


Fig. 2. Comparison of SD, MD-PD, DSD-PD for changing BER (with packet headers)

packet loss rate for each of the descriptions of the MD codec because of the smaller size of the packets when compared to G.729 packets, and the lower probability of losing both the descriptions simultaneously. DSD-PD performs the best for all BERs, but this better performance of DSD-PD comes at a penalty of additional bandwidth required to transmit two full rate streams at 16 kbps compared to 8 kbps required for SD and 9.2 kbps required for MD-PD. Hence, for a system with a bandwidth constraint, MD-PD is a good choice, since, for a small increase in bitrate (about 15%) compared to SD, we get a significant improvement in speech quality at high BERs.

*2) Packet headers:* Now we consider a more realistic scenario where headers are added to the speech packets by the lower protocol layers. In a typical 802.11 based ad-hoc network, headers would be added by RTP, UDP, IP and the 802.11 MAC layer protocol. The overheads for each packet add up to 68 bytes (the 802.11 MAC (28 bytes), IP (20 bytes), UDP (8 bytes) and RTP (12 bytes)), significantly larger than the payload which is a maximum of 10 bytes in our experiments. So the effective packet sizes are 78 bytes for G.729 and 75 or 73 bytes for the MD codec. For path diversity, the overheads are even larger because for each frame, we need to send 68 bytes of packet headers on both the paths. The difference in the payloads of the MD codec and the SD codec is insignificant now and the effective data rate of MD-PD is almost double that of SD. Note that sending duplicate copies of G.729 packets over two independent paths (DSD-PD) would require a bit rate of 124.8 kbps $((78+78)\times 8/10)$ while MD-PD needs 118.4 $((75+73)\times 8/10)$ kbps. For a small increase in required bandwidth, we can send two copies of G.729 packets instead of sending MD-G.729 packets that have only around half the information as a G.729 packet.

Figure 2 shows the performance of each of the methods for increasing BERs when packet headers are included. Note that there is a drop in the performance of all the methods compared to the no-header case of Fig. 1, because of larger $p$'s resulting from the larger packet sizes. We see that MD-PD performs better than SD but this gain in performance is achieved at a huge penalty in terms of the bandwidth required for transmission. The packet loss rate experienced by each description of the the MD codec is now almost the same as that of a G.729 packet because the ratio of their packet sizes is close to one. The better performance of MD-PD over SD can be attributed to better error concealment in MD-PD. When a packet is lost in only one of the paths, we need to conceal only about half of the bits in MD-PD, whereas, in the case of SD, no information is received if the single packet is lost in the network. Even after the inclusion of packet headers, DSD-PD performs significantly better than MD-PD (improvement in MOS by $0.72$ at a BER of $10^{-3.5}$) and this improvement in performance can be achieved at a small percentage increase (about 5.5%) in the number of bits transmitted.

TABLE III

PACKET SIZES WITH HEADERS

| Codec | Full headers (bytes) | Compressed Headers |
|---|---|---|
| G.729 | 78 | 40 |
| MD | 75 or 73 | 37 or 35 |

The best solution possible for improving packet efficiency is to use a header compression scheme to reduce the average size of the headers. RoHC (Robust Header Compression) is one such scheme that can be used to compress the IP/UDP/RTP headers to very small sizes of up to one byte. Efforts are underway to make RoHC compatible with 802.11 networks. The MAC layer header is still of significant size (28 bytes) and there are no MAC header compression methods at present. In Fig. 3 we plot the MOS values obtained for the new packet sizes with compressed headers. We assume that the IP/UDP/RTP headers are compressed to an average size of 2 bytes and the resultant packet sizes are listed in Table III. DSD-PD still provides the best quality of speech and requires only 10% more bits to be sent compared to MD-PD.

### C. Bursty Packet Losses

Next, we study the effect of bursty packet losses on each of the communication methods. We assume that burst losses are independent of packet sizes because they are usually caused due to phenomena like fading or shadowing in the network or

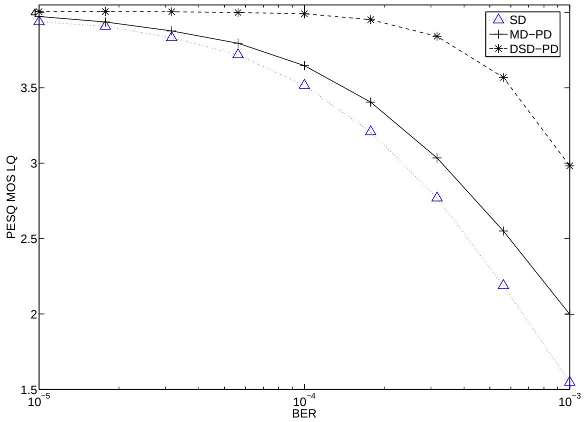


Fig. 3. Comparison of SD, MD-PD, DSD-PD for changing BER (with compressed packet headers)

other factors like a link failure. We model burst losses using the Gilbert model where the channel is modeled using a two-state Markov chain. The channel exists in either a good state or a bad state. No packets are dropped in the good state and all the packets are dropped when the channel is in the bad state. The same tracefiles were used for MD-PD and DSD-PD. Figure 4 shows the performance of each of the methods for an average burst size of 4 packets and different average packet loss rates. Observe that MD-PD performs significantly better than SD. This is because the packet loss concealment algorithms in CELP codecs are not very effective when successive packets are lost as the algorithm depends on the last received good frame to conceal the lost frame. Again, the DSD-PD method performs the best and the MOS delivered is significantly better than that of MD-PD. In a typical network with packet headers the advantage in performance provided by DSD-PD under burst loss conditions requires only a slight (5.5%) increase in the number of bits transmitted compared to MD-PD.

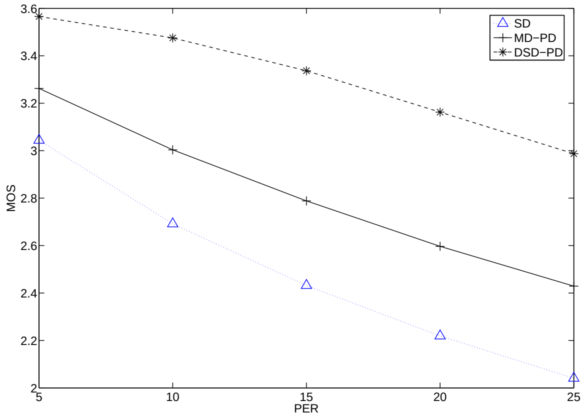


Fig. 4. Comparison under bursty packet losses only for av. burst size = 4

## IV. Conclusions

We consider the problem of supporting conversational voice in a wireless mobile ad-hoc network using path diversity. Among the three methods considered, we noticed that the simple path diversity method DSD-PD consistently performed better than the other two methods. Although this comparison might seem unreasonable since the source rate for DSD-PD is higher and almost double the source rate of both SD and MD-PD, we see that when the large packet headers are taken into consideration DSD-PD requires just 5% more bits compared to MD-PD. In a channel with uncorrelated bit errors, an MD codec has the advantage of having a smaller packet loss rate because of the smaller size of MD packets. But for low-rate speech codecs, this potential advantage of using a multiple description codec is mitigated by the large headers added at the various lower layer protocols. Also, note that the capacity of a network is significantly reduced when path diversity is used for voice communication. The small payloads compared to the large headers make any path diversity method highly inefficient and this inefficiency is difficult to overcome using a source coding diversity method like MD coding. In Table IV we show the ratio of the number of bits transmitted for each method compared to SD.

TABLE IV
Ratio of number of bits transmitted for each method

| Method | without headers | with headers |
|---|---|---|
| SD | 1 | 1 |
| MD-PD | 1.15 | 1.90 |
| DSD-PD | 2 | 2 |

For ad-hoc networks based on 802.11, we do not see a compelling reason to adopt multiple description coding for speech. Simple path diversity using a good quality low-rate speech codec like G.729 is a better option considering the quality of speech delivered by these codecs. If, in the future, the packet headers are compressed to a size smaller than the speech packets, then using a multiple description codec may be beneficial in terms of bandwidth and quality.

## References

[1] H. Dong, I. Chakeres, H. Dong, E. Belding-Royer, A. Gersho, and J. Gibson, "Selective bit-error checking at the mac layer for voice over mobile ad hoc networks with ieee 802.11," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC)*, Mar. 2004.

[2] A. Servetti and J. D. Martin, "Link-level unequal error detection for speech transmission over 802.11 networks," in *Proc. Special Workshop in Maui (SWIM)*, Jan. 2004.

[3] M. Petracca, A. Servetti, and J. D. Martin, "Voice transmission over 802.11 wireless networks using analysis-by-synthesis packet classification," in *First International Symposium on Control, Communications and Signal Processing*, 2004.

[4] A. Servetti and J. D. Martin, "Adaptive interactive speech transmission over 802.11 wireless lans," in *Proc. IEEE Int. Workshop on DSP in mobile and Vehicular Systems*, Apr. 2003.

[5] H. Dong, A. Gersho, V. Cuperman, and J. Gibson, "A multiple description speech coder based on amr-wb for mobile ad hoc networks," in *ICASSP'04*, May 2004.

[6] *Coding of Speech at 8 kbit/s Using Conjugate- Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)*, ITU-T Recommendation G.729 Std.

[7] A. W. Rix, "Comparison between subjective listening quality and P.862 PESQ score," in *Proceedings of Online Workshop Measurement of Speech and Audio Quality in Networks,*, May 2003.